Employing parent-offspring analysis to estimate the proportion of hatchery-reared fish in fisheries

Per J. Palsbøll

Dept. of Environmental Science, Policy and Management

University of California at Berkeley

Other kinds of "fishes"

- Genetic tagging of
  - individuals
    - Abundance estimation by mark-recapture
    - Wildlife DNA Registers of legal harvest
  - close relatives
    - Abundance
    - Operational sex ratio
    - "real-time' migration
    - Validation of demographic models

Objective?

- Estimate of the proportion of hatchery-reared salmon in fisheries?
  - Desirable characteristics
    - Accuracy
    - Rapid turnover in terms of estimating the proportion of hatchery-reared fish
    - Low cost

"Population-level" approaches:

- Proportions of each breeding population, from
  - population allele frequencies
  - assignment of sampled individuals
- Requirements
  - Sample all putative source populations
  - Some degree of divergence among source populations
- Violation of model assumptions
  - Panmixis
  - Temporal differences in allele frequencies

Individual-based approaches: Individual identification

- Genetically "tag" all released individuals
- Match samples from fisheries to an individual-based DNA register
  - Norwegian minke whale DNA register

- While feasible in a few 1000s samples, not the case with hatchery-reared salmon where many more individuals are released

Individual-based approaches: Assigning offspring to parents

- *All* hatchery-reared fish from hatchery breeding programs will match at one or two alleles at each locus to one breeding stock female and one breeding stock male
  - *All* hatchery-reared individuals are genetically labeled
  - *All* can be identified with a high degree of confidence given sufficient genetic markers

Tiger prawns, *Panaeus esculentus*

- Bravington MV and Ward RD (2004) Microsatellite DNA markers: evaluating their potential for estimating the proportion of hatchery-reared offspring in a stock enhancement programme. *Molecular Ecology* **13**:1287-1297

All hatchery-incompatible individuals from natural populations

$$p_{inc} = (1 - p_H) \times (1 - p_{CbC})$$

- p_H = proportion of fisheries from hatchery
- P_{inc} = proportion of individuals that is incompatible with a hatchery origin
- p_{CbC} = proportion of non-hatchery individuals compatible with a hatchery origin by chance

Fraction in fisheries of hatchery origin

$$\hat{p}_H = 1 - \frac{n_{inc}}{n\left(1 - p_{CbC}\right)}$$

- $P(hat)_H$ = proportion of fisheries from hatchery

- n = total number of individuals sampled in the fisheries

- n_{inc} = number of individuals in the fisheries incompatible with a hatchery origin

- p_{CbC} = proportion of non-hatchery individuals compatible with a hatchery origin by chance

How is p_{CbC} estimated?

- Population allele frequencies estimated from empirical data
  - "Standard" assumptions
    - Random mating
    - Independent segregating loci
  - In this case conditioned upon having at least one allele in common with the sampled "offspring" at each locus and the alleles present in the hatchery breeding stock

Empirical data (8 STR loci)

Locus	Repeat type	Sample size	Alleles	H_E	P_I	$p_{ChC,DS}$	$p_{ChC,D}$
CSGES120	Tri	183	11	0.496	0.290	0.538	0.865
CSGES189	Tri	186	11	0.737	0.112	0.309	0.669
CSGES047	Tetra	187	19	0.889	0.022	0.076	0.363
CSGES176	Tri	186	15	0.843	0.040	0.126	0.466
CSGES268	Tri	183	16	0.713	0.099	0.239	0.652
CSGES190	Tri	180	8	0.518	0.270	0.523	0.854
Pe1.1	Di	186	27	0.914	0.013	0.047	0.292
Pmcd01	Tri	185	10	0.828	0.051	0.163	0.512
Overall					5.1×10^{-10}	1.5×10^{-6}	8.2×10^{-3}

Table 4 Variability parameters for eight microsatellite loci in *Penaeus esculentus*

H_E is Hardy–Weinberg expected heterozygosity, P_I is probability of identity of two randomly picked animals in the wild population (Paetkau & Strobeck 1994) and is included for comparison, $p_{ChC,DS}$ is the probability that a randomly-chosen female-male pair is compatible with being the parents of an unrelated animal, and $p_{ChC,D}$ is the probability that a randomly-chosen female is compatible with being the mother of an unrelated animal.

Expectations with 8 STR loci

Table 5 Per cent CV of estimated p_H

Brood pairs	Captures	$p_H = 0.01$			$p_H = 0.03$			$p_H = 0.10$			$p_H = 0.30$		
		D	D(S)	DS	D	D(S)	DS	D	D(S)	DS	D	D(S)	DS
1	10	*	*	*	*	*	*	99	95	95	49	48	48
1	100	*	100	100	64	57	57	31	30	30	**15**	**15**	**15**
1	1000	42	31	31	**20**	**18**	**18**	**10**	**9**	**9**	**5**	**5**	**5**
5	10	*	*	*	*	*	*	*	95	95	52	48	48
5	100	*	100	100	88	57	57	36	30	30	**16**	**15**	**15**
5	1000	72	31	31	28	**18**	**18**	**11**	**9**	**9**	**5**	**5**	**5**
10	10	*	*	*	*	*	*	*	95	95	55	48	48
10	100	*	100	100	*	57	57	41	30	30	**17**	**15**	**15**
10	1000	97	32	31	35	**18**	**18**	**13**	**9**	**9**	**5**	**5**	**5**
50	10	*	*	*	*	*	*	*	95	95	79	48	48
50	100	*	100	100	*	57	57	74	30	30	25	**15**	**15**
50	1000	*	32	32	76	**18**	**18**	23	**9**	**9**	**8**	**5**	**5**
100	10	*	*	*	*	*	*	*	95	95	*	48	48
100	100	*	*	100	*	57	57	*	30	30	35	**15**	**15**
100	1000	*	32	32	*	**18**	**18**	35	**10**	**9**	**11**	**5**	**5**
1000	10	*	*	*	*	*	*	*	96	96	*	49	48
1000	100	*	*	*	*	60	58	*	30	30	*	**15**	**15**
1000	1000	*	36	34	*	**19**	**18**	*	**10**	**10**	*	**5**	**5**

Bold figures are below 25%, asterisks exceed 100% (see comments). D indicates dam-only typing, D(S) indicates dam and partly typed sire, DS indicates dam and fully typed sire.

Conclusions

- At this level of genotypic data (8 STR loci) the amount of data is the main issue:
  - High CV when
    - proportion of hatchery-reared fish is low in fisheries,
    - sample size from fisheries is low,
    - only female brood stock genotypes, and
    - (presumably) if natural populations are poorly sampled

$$\hat{p}_H = 1 - \frac{n_{inc}}{n\left(1 - p_{CbC}\right)}$$

Adding loci improves precision

Table 6 Per cent CV of estimated p_H obtainable with three extra loci; dam-only typing

Brood pairs	Captures	$p_H = 0.01$ D	$p_H = 0.03$ D	$p_H = 0.10$ D
5	10	*	68	32
5	1000	47	**21**	**10**
10	10	*	77	34
10	1000	59	24	**11**
20	10	*	93	37
20	1000	77	29	**12**

Bold figures are below 25%, asterisks exceed 100% (see comments).

Experimental design issues: Hatchery-end

- Decrease p_{CbC}
  - Complete genotypes for all brood stock individuals
    - Bravington and Ward (2004)
  - All breeding combinations known
  - Increase the number and/or variability of genetic markers typed in each specimen

- Size of p_{CbC} likely to depend upon source of hatchery breeding stock
  - i.e., from natural source or a maintained separate breeding stock

$$\hat{p}_H = 1 - \frac{n_{inc}}{n(1 - p_{CbC})}$$

Experimental design issues: Natural populations

- Estimate p_{CbC} as accurately as possible

  increasing sample sizes and/or employing loci with fewer alleles per locus

  Dealing with temporal fluctuations in allele frequencies

- Reduce p_{CbC} as much as possible

  Increase the number and/or variability of the genetic markers typed in each individual

$$\hat{p}_H = 1 - \frac{n_{inc}}{n\left(1 - p_{CbC}\right)}$$

Experimental design issues: Fisheries

- Sampling
  - The scale of sampling from the fisheries need be adjusted to the expected proportion of hatchery-reared fish in order to obtain the appropriate precision

$$\hat{p}_H = 1 - \frac{n_{inc}}{n\left(1 - p_{CbC}\right)}$$

Experimental design issues: Rapid, reliable and cost effective data

- Genotyping errors
  - Employ excess of number of genetic marker
  - Estimate error rates
- Null-alleles
  - If detected redesign primers to rid null-alleles
- Rapid and reliable genotyping
  - Employ SNPs as opposed to STRs
    - Easily automated in terms of data generation and scoring
    - No inter-laboratory standardization issue
    - ~23+ SNPs with frequencies of .4/.6 have the same probability of identity as the 8 STR loci in Bravington & Ward (2004)

File   View

SNP genotyping
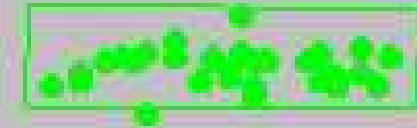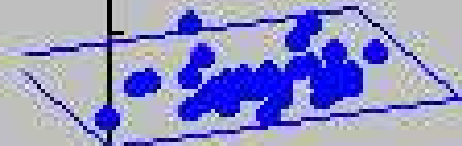
- Easy to automate all steps
- Small target regions, applicable to degraded DNA
- Hardware relatively inexpensive
- New technologies developing at a rapid rate

C10

C

A

Experimental design issues: Data analysis

- Target value for p_{CbC}
  - Standard panmictic model, unrelated individuals?
  - Stock enhancement;
    - Hatchery breeding stock individuals may have close relative in natural population

      Use conservative cutoff for p_{CbC} (e.g., full siblings to breeding stock males and females)

$$\hat{p}_H = 1 - \frac{n_{inc}}{n(1 - p_{CbC})}$$

Thank You